A fully Automated CAD system using Multi-category Feature Selection with Restricted Recombination

Ranadhir Ghosh, Moumita Ghosh, John Yearwood, Subhasis Mukherjee
University Of Ballarat
{r.ghosh, m.ghosh, j.yearwood}@ballarat.edu.au

Abstract

In pattern recognition problems features plays an important role for classification results. It is very important which features are used and how many features are used for the classification process. Most of the real life classification problem uses different category of features. It is desirable to find the optimal combination of features that improves the performance of the classifier. There exists different selection framework that selects the features. Mostly do not incorporate the impact of one category of features on another. Even if they incorporate, they produce conflict between the categories. In this paper we proposed a restricted crossover selection framework which incorporate the impact of different categories on each other, as well as it restricts the search within the category which searching in the global region of the search space. The results obtained by the proposed framework are promising.

1. Introduction

Computer Aided Diagnosis (CAD) is a field in pattern recognition dealing with medical images for detection and classification of Microcalcifications. CAD's main aim is to provide doctors with an accurate and reliable second opinion by analysing a digital image.

In the literature, various numbers of techniques are used to classify the presence of microcalcifications in digital mammograms. Those are wavelet-based techniques [1], statistical techniques [2], image processing techniques [3-4], neural networks [5-11] or a combination of techniques [3].

In practical pattern recognition problems, a classification function learns through an inductive learning algorithm that maps a given input pattern to one of the existing classes of the systems. However the classifier can work well when a meaningful set of input feature is provided to it. Only a particular type of feature such as statistical or structural alone may not be the best possible choice. Hence a combination of different categories of features from the raw data set might provide very useful information for the classifier. This type of combination however leads to the formulation of multi category features as input set. In an addition the length of the feature vector thus increases to many extent. It has been observed that, beyond a certain point, the inclusion of additional features leads to a worse rather than better performance. Moreover, the choice of features to represent the patterns affects several aspects of pattern recognition problem such as accuracy, required learning time and necessary number of samples. Therefore the main goal of feature subset selection is to reduce the number of features used in the classification while maintaining acceptable classification accuracy.

In general, feature selection algorithms have two components: an evaluation function that scores candidate feature sets, and a search engine for finding those sets. Given a set of features the selection algorithm will examine a series of sets of features, and choose the one that maximizes the evaluation function. Recent comparative studies of feature selection algorithms can be found in [12], [13].

A common problem with the multi category feature classification is the conflict between the categories. None of the feasible solutions allow simultaneous optimal solution for all categories. Whether an optimal solution for all categories leads to an optimal solution for one combined set of mixed multicategory feature can be another research question.

Ghosh et al [15] have proposed a multi –category framework for Brest cancer recognition, which was capable of handling multi category features for the classifier. The method divides different set of features into different category and pass those through different selection modules. The main problem with that structre is that there is no any interconnection between the lselection modules. Hnece we can not measure the

6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007)
0-7695-2841-4/07 $25.00 © 2007 IEEE

impact of one set of features on other.It has been observesd that the individual selected category could work well, but when they are combined together they lead to dcerase in classification accuracy. Whice leads to the conclusion of using ingle selection module instad. But single selection module makes the recombination more complicated. Crossover combines two parent chromosomes to produce a new offspring. The idea behind crossover is that the new chromosome may be better than both of the parents if it takes the best characteristics from each of the parents. In general selection, each category of features will be treated uniformly. For single characteristics or category this will not cause any problem. But for multiple characteristics of feature, different characteristics will be combined together to produce the offspring. There could be a chance have mix offspring in next generation that can mislead the results. Which insists to formulate the modular architecture.

In order to overcome this problem we propose another modular architecture that modularize the selection procedure during the recombination and combine the selection procedure during evaluation of fitness. In order to find an optimal solution, the search space is divided based on an individual category in each sub region in times of offspring generation the and but in times of selection and evaluation we take the whole search space as a single one. Hence instead of having different selection modules, we have now different recombination module with a single selection module. The recombination modules work independently. Even though we uses a single selection module, but we are dividing the serach space into different sub regions in times of recombination hence the time for the propsed selection to reach optimal solution will be much faster than general slelection where the recombination is doen in the single search space. A good parallel implementation of the algorithm can have a much better time complexity than the general selection method.

2. Research methodology

The research methodology can broadly be classified into four modules, such as Preprocessing, Area Clustering, Feature extraction, Feature subset selection, Neural network based classifier.

2.1. Preprocessing

Digital mammogram database is taken from university of South Florida [18]. Each case contains four mammograms from a screening exam. Once digital mammogram decompressed, suspicious area extracted from the mammogram. Suspicious area is already marked in all digital mammograms of DDSM by three expert radiologists. Starting position of suspicious area and chain code value for suspicious area extraction are available from *". OVERLAY"* file. DDSM is available in form of cases. Each case contains four mammograms from a screening exam.

2.2. Area clustering

The next step is to choose the suspicious from the image to perform the classification task. The gray level of each pixel after preprocessing has been considered in this case for describing the image. We cluster the mammograms into different suspicion region with the gary level properties of each pixel. The suspicion regions differ from the normal region in terms of average histograms of those areas. We calculate the average grey level that gives an indication of the grey level value associated with the suspicious area and the normal area. Suppose $S = \{s_i, i = 1,.....,m\}$ is the set of suspicious areas in the mammogram I, where s_i's are the suspicious areas, and there are m number of such areas in I. The average grey level at the suspicious are is the summation of the grey level value for all the pixel in that are divide by the total number of pixel.

Hence average grey level of the suspicious are s_i is given as follows $AG(s_i) = a_i \frac{\sum_{j=1}^{n} g_{ij}}{n_i}$, where g_{ij} denotes the grey level value in I at pixel j of ith suspicious area, and n_i denotes the total number of pixel in that area , and a_i is a constant.

We also calculate the average grey level value around the outside of the suspicious are. We define a fixed width of strip around suspicious area, and calculate average grey level value of that strip. We call it Outer Average grey density. Hence if there is m number of suspicious area then there will be m number of outer boundary around each suspicious area. Let $O = \{o_i, i = 1,.....,m\}$ is the set of outer boundaries in the mammogram I. The Outer Average grey density of the outer region o_i is given as follows $OAG(o_i) = b_i \frac{\sum_{k=1}^{p} g_{ik}}{p}$, where g_{ik} denotes the grey level value in I at pixel k of ith outer region, and b_i is a constant.

Our main aim is to maximize the difference of the Average grey density and the Outer Average grey density for each suspicious area.

Hence the objective function for a particular suspicious area s_i is as follows $f(s_i) = \max(abs(AG(s_i) - OAG(o_i)))$

If there are m number of suspicious area then the overall objective of the clustering algorithm is to minimize the summation of $f(s_i)$ for all s_i. Hence the objective function is as follows

$$f_1 = \max\left(\sum_{i=1}^{m}(abs(AG(s_i) - OAG(o_i)))\right). \quad (1)$$

If we write in terms of minimization the objective function becomes

$$f_1 = \min\left(-\sum_{i=1}^{m}(abs(AG(s_i) - OAG(o_i)))\right) \quad (2)$$

We also want to restrict the area of the suspicious region to certain level. If the suspicious area is too big compare to the outer region or the actual image, then there would be a chance of mixing normal area with the damaged area. Hence we introduce the second objective to minimize the size of the suspicious area. The second objective function is as follows

$f_2 = \min\left(\frac{\sum_{i=1}^{m} n_i}{T}\right)$, where n_i denotes total number of pixel in i[th] suspicious area, and T denotes total number of pixel in the mammogram I. Hence the actual objective function for the clustering algorithm is as follows

$$f = \min\left(\lambda \frac{\sum_{i=1}^{m} n_i}{T} - \sum_{i=1}^{m}(abs(AG(s_i) - OAG(o_i)))\right) \quad (3)$$

2.3. Clustering method

In this section we describe an algorithm for solving cluster analysis problem. Algorithm [17] contains some steps, which deserve some explanations. In Step 1 the center of the entire set A is calculated with respect to a given norm. In this case the problem is a convex programming problem. In Step 2 we calculate a center of the next (k+1)-th cluster, assuming the previous k cluster centers to be known and fixed. It should be noted that the number of variables in the problem is n, which is substantially less than if we calculate all cluster centers simultaneously. In Step 3 the refinement of all k+1 cluster centers is carried out. One can expect that the starting point $x^{k+1,0}$ calculated in Step 2 is not far from the solution to the problem. Therefore it takes only a moderate number of iterations to calculate it. Such an approach allows one to significantly reduce the computational time for solving the problem.

One of the important questions when one tries to apply Algorithm 1 is the choice of the tolerance $\epsilon > 0$. Large values of ϵ can result in the appearance of large clusters whereas small values can produce small and artificial clusters. Appropriate values for ϵ are $\varepsilon \in [10^{-1}, 10^{-2}]$. The objective function in the clustering problem is Lipschitz continuous. Discrete gradient method is applied to solve the problems. The discrete gradient is a finite difference estimate to a subgradient. Unlike many other finite difference estimates to subgradient, the discrete gradient is defined with respect to a given direction, which allows a good approximation for the quasidifferential. The algorithm calculates discrete gradients step by step, and after a finite number of iterations either the descent direction is calculated or it is found that the current point is an approximate stationary point. In the Discrete gradient method Armijo's algorithm is used for a line search. Hence at a given approximation, the method calculates the descent direction by calculating the discrete gradients step by step, and improving the approximation of the Demayibv-Rubinov quasidifferential. Once the descent direction is calculated, Armijo's algorithm is used for line search. The local minimum is chosen as the next approximation. Hence the Discrete Gradient method jumps over many local minima and finds very deep local minima.

2.4. Feature extraction

Feature found in literature are structural based features, statistical based features, and spatial grey level dependence matrix based features. These features are found in [16].

2.5. Feature selection

In this section we explain the selection mechanism. We are comparing our proposed module with other two existing selection mechanism. Those are the modular structure, and the general structure respectively. We have explained those structures in details later in this section.

2.5.1 Restricted Crossover selection Architecture*:* There exists one single selection module. The selection module contains different recombination modules those are responsible to create the offspring for the next generation for a single category of feature. The selection module evaluates the fitness with respect to

all the categories. Then it creates the next generation through the roulette wheel selection mechanism. After selecting the features for all categories it distributes different categories of feature among the recombination modules. The recombination module then apply crossover for the individual category of features. After the recombination we merge the categories those together and pass it to the selection module. Hence features are categorized in recombination phase. After feature selection is done the selected features are passed through the decision system. The decision system is a Nural Network that is responsible for classifying the input.

2.5.2 Modular selection Architecture: We are using this architecture to compare the how the prosed alrotithm has imporeved. Each of the modules works independently on its own domain. They are built and trained for its specific task. Each of them is responsible to find out the best combination of features from each category. Hence the features are categorized since the beginning. The final decision is made on the results of the individual networks, often called expert networks or agents. Here also the decision system is a Nural Network that is responsible for classifying the input.

2.5.3 General selection Architecture*:* We are using this architecture to compare the how the prosed alrotithm has imporeved. There is one selection module that contains one recombination module only. The features re passed through the selection module which evaluates the fitness and generates the next generation applying crossover and selection operator. Hence all the featres are treated together, they are not categorized at all.

2.6 Selection module

As described earlier, selection module is responsible to select the best combination from a given set of features as input. Feature selection algorithms have two components: an evaluation function that scores candidate feature sets, and a search engine for finding those sets. The training phase and the evaluation phase work together. The chromosome mask will be different for different category of features. In the evaluation phase the population is initialized randomly. For each member in the population, if the bit position holds a zero value the feature is assigned to zero and a new data set is created. With that dataset the neural network is trained. So, for individual members in the population, there are individual neural networks that have to be trained with the separate dataset. The traditional EBP algorithm is used to train the neural network, then that trained neural network is used to calculate the fitness. To calculate the fitness of the individual population, the feature vector is multiplied by the individual population. If a particular feature is not selected, that place is neutralized. So the feature is multiplied by zero and deactivates its effect on fitness.

In Restricted Crossover selection mechanism there is one selection module, which takes all categories of feature. The dataset is passed through this selection module. The selection module initializes the individuals in the population randomly with the value 1 or 0. After initialization, the selection module modifies the dataset accordingly (1 means that feature is present and 0 means it is not present in the dataset) for individuals in the population. Then the neural network is trained for the individuals in the population. The training classification accuracies of the individual are taken as the fitness. Then the roulette wheel selection mechanism is applied to generate the next generation. Then the features are categorized into different category and passed the categorized individuals into different recombination module. The recombination modules apply crossover that is restricted within the category. Once the recombinations are done, the individuals are merged again and passed through the next generation. The stopping criteria of the selection module are number of generation and the training classification accuracy of the neural network.

In modular selection mechanism, the features are categorized at the beginning into different category and the dataset are form for individual category. There exist as many selection modules as the number of category of feature. Each category of features is passed through the individual selection mechanism. Those work independently. The stopping condition for training the neural network is to be the same for all the members in the population and it is taken as the classification error.

In general selection there is only one selection module that does the recombination as well. There is no any categorization. The features are treated uniformly. Hence during the crossover, there is no restriction and any feature can recombine with other to form the next generation.

2.7 Decision module

A decision module is responsible for classifying the results on the basis of output of each selection module. We use a Neural Network as a decision system. Output form each selection module is fed to the decision NN. Depending on the feature selected from the different selection module, the decision neural network classifies the input pattern in three classes (Malignant, Benign, Normal).

2.8 Evolutionary strategy based classifier module

Evolutionary algorithms (EAs) are search methods that take their inspiration from natural selection and survival of the fittest in the biological world. EAs differ from more traditional optimization techniques in that they involve a search from a "population" of solutions, not from a single point. Each iteration of an EA involves a competitive selection that rejects the poor solutions. The solutions with high "fitness" are "recombined" with other solutions by swapping parts of a solution with another. Solutions are also "mutated" by making a small change to a single element of the solution. Recombination and mutation are used to generate new solutions that are biased towards regions of the space for which good solutions have already been seen.

3. Experimental results

3.1 Dataset

The proposed approach has been implemented in C++ and UNIX. We have used 100 cases [14] of each Malignant, Benign and Normal for training. Hence the length of the training dataset was 300. Also we have used 20 cases of each Malignant, Benign and Normal for testing. Hence the length of the testing dataset was 60.

3.2 Clustering Results

We compare the area clustering result with the area extracted by the doctors. Altogether there are 100 damaged areas. Our clustering algorithm has detected most of the damaged areas and also has detected few areas that look like a microcalcification. The clustering results are shown in Table I.

TABLE I CLUSTERING RESULTS

Number of Damaged Area			
By Doctor	By Clustering Algorithm		
	Correct	Missed	Wrong
100	98	2	40

Table1 shows that our clustering algorithm has missed only 2% of the damaged area and detected 98% correctly. It has also detected 40 non-damaged areas those are looked like micro-calcification. The classification algorithm is responsible to classify it as normal.

3.3 Classification Results

The experimental results are shown in Table II. The results obtained by the proposed framework are compared with the Modular and General framework respectively. Table II shows the training classification accuracy for the three frameworks. The 'M' stands for Malignant, 'B' stands for Benign and 'N' stands for Normal cases.

TABLE ii classification results

Model	Training results (Classification accuracy %)			Testing results (Classification results %)		
	M	B	N	Type1 error	TypeII error	Total accuracy
Restricted	93	90	92	3.33	6.66	90
Modular	91	90	88	3.33	8.33	88.33
General	80	70	70	6.66	20	73.33

The following table (Table III) shows the shows the time take for each algorithm to converge to a solution.

TABLE III Computational time results

Model	Training Time (minutes)
Restricted	135
Modular	110
General	140

The above table (Table III) shows that the time take for the Modular Network is much lesser than General framework and the proposed Restricted Crossover Framework. This is because the Modular framework has several selection modules, which are run parallely. The restricted framework works bit faster than the general framework because the restricted framework has got different recombination module, which are run parallely. Hence in terms of computational time the Modular framework is better. But there is a clear trade off between quality of solution and the time taken to converge. For our case the quality of the solution is much more important that the converging time. And the classification accuracy on testing dataset for the proposed method is 90%, for modular framework is 88.33% and for general framework is 73.33%. Hence the proposed algorithm is much better that the earlier attempts in terms of quality of solution.

4. Conclusion

In this paper we have presented a fully automated breast cancer recognition system. The system is capable of clustering the suspicious region in the mammogram by itself. The result is compared with the experts and the result shows promising. The paper

proposes a restricted crossover selection framework that is suitable for multi category feature selection. The selection module uses combination of GA and neural network classifier. We have tested with Digital Mammogram dataset. We have used 40 features divided into three category of statistical, structural and dependency features. We got 90% test classification accuracy. We have shown the improvement of classification accuracy on both the training and testing dataset over the Modular selection framework and the general selection framework.

REFERENCES

1. Yoshida H, Nishikawa R, Maryellen G and Doi K, Computer-aided Diagnosis in Mammography: Detection of Clustered Microcalcifications Based on Multiscale Edge Representation, Computer Assisted Radiology, 1996, Elsevier Science.
2. Solka J, Poston W, Priebe C, Rogers G, Richard L, Marchette D, Woods K and Bowyer K, The Detection of Microcalcifications in Mammographic Images Using High Dimensional Features, Seventh Annual IEEE Symposium on Computer-Based Medical Systems, 1993, pp. 139-145.
3. Shen L, Rangayyan R and Desaultels J, Detection and Classification Mammographic Calcifications, International Journal of Pattern Recognition and Artificial Intelligence, 1994, World Scientific Publishing, pp. 1403-1416.
4. Barman H, Granlund G and Haglund L, Feature Extraction for Computer-Aided Analysis of Mammograms *in State In The Art of Digital Mammographic Image Analysis*, 1994, World Scientific Publishing, 1994, Vol. 7, No. 6, pp. 128-147.
5. Chitre Y, Dhawan A and Moskowtz M, Artificial Neural Network Based Classification of Mammographic Microcalcifications Using Image Structure Features *in State In The Art of Digital Mammographic Image Analysis*, 1994, World Scientific Publishing, Vol. 7, No. 6, pp. 167-197.
6. Zheng B, Qian W and Clarke L, Multistage Neural Network for Pattern Recognition in Mammogram Screening, IEEE International Conference on Neural Networks (ICNN), 1994, pp. 3437-3447.
7. Woods K, Doss C, Bowyer K, Solka J, Priebe C and Kegelmeyer W, Comparative Evaluation of Pattern Recognition Techniques For Detection of Microcalcifications in Mammography, International Journal of Pattern Recognition and Artificial Intelligence, 1994, World Scientific Publishing, Vol. 7, No. 6, pp. 1417-1436.
8. Z. Huo, M. Giger, C. Vyborny, D. Wolverton, R. Schmidt, K. Doi, Computer-Aided Diagnosis: Automated Classification of Mammographic Mass Lesions, *Proceedings of the 3rd Int. Workshop on Digital Mammograph*, Chicago, IL, pp. 207-211, 1996.
9. H. Chan, D. Wei, M. Helvie, B. Sahiner, D. Adler, M. Goodsitt., and N. Petrick, Computer-Aided Classification of Mammographic Masses and Normal Tissue: Linear Discriminant analysis in Texture Feature Space, *Phys. Med. Biol.*, vol.40, pp. 857-876, 1995.
10. D. Meersma, P. Scheunders , and D. Van Dyck, Detection of Microcalcifications using Neural Networks, *Proceedings of the 3rd Int. Workshop on Digital Mammograph, Chicago*, IL ,pp. 97-103, 1996.
11. A. P. Dhawan, Y. Chite, C. Bonasso, and K. Wheeler, Radial-Basis-Function Based Classification of Mammographic Microcalcifications Using Texture Features, *IEEE Engineering in Medicine and Biology & CMBEC*, pp. 535-536, 1995.
12. A. K. Jain, and D. Zongker, Feature selection: evaluation, application and small sample performance, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 19, no. 2, pp. 153 – 158, 1997.
13. R. Kohanvi, and G. H. John, Wrappers for feature subset selection, *Artificial Intelligence*, vol. 97, pp. 273 – 324, 1997.
14. R. Ghosh, M. Ghosh, J. Yearwood, A Modular Framework for Multi category feature selection in Breast Cancer recognition, *Accepted in 12th European Symposium on Artificial Neural Networks,* 2004.
15. R. Ghosh, M. Ghosh, J. Yearwood, A hybrid evolutionary algorithm for multi category feature selection in Breast Cancer recognition, Accepted 2nd International Conference on Soft Computing and Intelligent Systems (SCIS), 2004.
16. B. Verma, J. Zakos, A Computer-Aided Diagnosis System For DigitalMammograms Based On Fuzzy-Neural And Feature Extraction Techniques, IEEE Transactions on Information Technology in Biomedicine, vol 5, pp.46-54, 2001.
17. A.M. Bagirov. Derivative-free methods for unconstrained nonsmooth optimization and its numerical analysis. Investigacao Operacional, pp. 19-75, 1999.
18. ftp://figment.csee.usf.edu.au.

0-7695-2841-4/07 $25.00 © 2007 IEEE